(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 3 243 165 B1

(12) EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention of the grant of the patent:
02.08.2023 Bulletin 2023/31

(21) Application number: **15877226.9**

(22) Date of filing: **21.12.2015**

(51) International Patent Classification (IPC):
G06F 18/25 (2023.01) ***G06V 10/80*** (2022.01)
G06V 20/52 (2022.01) ***G06V 20/54*** (2022.01)

(52) Cooperative Patent Classification (CPC):
G06V 20/54; G06F 18/256; G06V 10/811; G06V 20/52; G06V 2201/08

(86) International application number:
PCT/SE2015/051382

(87) International publication number:
WO 2016/111640 (14.07.2016 Gazette 2016/28)

(54) METHOD AND SYSTEM FOR CATEGORIZATION OF A SCENE

VERFAHREN UND SYSTEM ZUR KATEGORISIERUNG EINER SZENE

PROCÉDÉ ET SYSTÈME DE CATÉGORISATION D'UNE SCÈNE

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priority: **07.01.2015 SE 1550006**

(43) Date of publication of application:
15.11.2017 Bulletin 2017/46

(73) Proprietor: **Viscando AB**
417 55 Göteborg (SE)

(72) Inventors:
- **SINGH, Amritpal**
  423 49 Torslanda (SE)
- **ERLANDSSON, Ulf**
  439 91 Onsala (SE)

(74) Representative: **Kransell & Wennborg KB**
P.O. Box 27834
115 93 Stockholm (SE)

(56) References cited:
EP-A2- 0 955 606
WO-A1-2006/082429
WO-A1-2010/124062
WO-A1-2012/029058
WO-A1-2013/029674
CN-A- 101 388 145
CN-A- 103 324 955
JP-A- 2005 215 909
US-A- 5 724 447
US-A1- 2005 105 765
US-A1- 2005 201 612
US-A1- 2006 227 862
US-A1- 2007 122 040
US-A1- 2007 237 398
US-A1- 2009 010 490
US-A1- 2009 010 490
US-A1- 2009 195 371
US-A1- 2010 092 078
US-A1- 2010 194 883
US-A1- 2012 087 572
US-A1- 2012 115 598
US-A1- 2012 140 061
US-A1- 2013 182 114
US-A1- 2014 339 374
US-B1- 6 570 608

- **BOTA S ET AL: "A framework for object detection, tracking and classification in urban traffic scenarios using stereovision", INTELLIGENT COMPUTER COMMUNICATION AND PROCESSING, 2009. ICCP 2009. IEEE 5TH INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 27 August 2009 (2009-08-27), pages 153-156, XP031545115, ISBN: 978-1-4244-5007-7**

EP 3 243 165 B1

Description

TECHNICAL FIELD

[0001] The present invention relates to a method for categorizing a moving object within a scene. In particular, the present invention relates to a method specifically taking into account the three dimensional data within a scene for determining the location of and type of objects present within the scene. The invention also relates to a corresponding system and a computer program product.

BACKGROUND OF THE INVENTION

[0002] Camera systems are normally used within the security business for surveying a scene. The low cost of the cameras used and the simplicity involved for communicating with the cameras, typically over the Internet, provides for further possible applications such for example for use within a shop for determining a customer shopping behavior.

[0003] However, the collection of video data may unfortunately in some instances potentially compromise the privacy of anyone observed by the system. There is thus a trade-off between the advantages made possible with video analysis and the thereto related privacy issues. This is specifically apparent within an everyday situation where a person is typically not conscious of the fact that video capture is taking place, such as within the above mentioned shop, within a park, etc.

[0004] A solution to this problem is provided in US 7,602,944, disclosing a computer-implemented method of identifying a moving object in a digital video stream of a scene. The identification of the moving object is achieved by threshold subtracting a current video frame from a short term average video frame. The video frame is typically of an in comparison low resolution, i.e. reduced resolution, whereby generally all details will be discarded and the privacy for e.g. the person(s) captured by the video camera is upheld. The objects detected within the scene may be "followed" through a plurality of subsequent video frames, allowing the direction of the object to be determined as well as for determining the number of objects passing into and out from a defined zone within the scene.

[0005] Even though the disclosure of US 7,602,944 provides some relief to the privacy issues discussed above, there will be problems as the robustness of the system, specifically in relation to scenes with varying illumination, such as for example in an outdoor scenario. Accordingly, there is a need for a more robust method for object detection within a scene, preferably also allowing for the possibility of extracting further information relating to the detected objects while still keeping the privacy at an in comparison high level.

[0006] Further attention is drawn to US20130182114. US20130182114 makes use of a video camera and a depth sensor for determining a real life height and width of an object. In an organized environment, such as within a retail space, the method disclosed in US20130182114 may be used for object separation and subsequent tracking of the objects.

[0007] In addition, CN101388145 provides an automatic warning method for improving road safety. The disclosed method uses single camera "two-dimensional" object motion detection, including determination of a size and a speed of the object.

[0008] Further attention is drawn to BOTA S ET AL: "A framework for object detection, tracking and classification in urban traffic scenarios using stereovision, INTELLIGENT COMPUTER COMMUNICATION AND PROCESSING, 2009. ICCP 2009. IEEE 5TH INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 27 August 2009 (2009-08-27), pages 153-156, XP031545115,ISBN: 978-1-4244-5007-7, disclosing a framework for obstacle detection.

[0009] In addition, US20090195371 presents a method and apparatus for performing collision detection. Furthermore, US2009010490 presents a method of identifying, tracking, and counting human objects of interest based upon at least one pair of stereo image frames taken by at least one image capturing device.

SUMMARY OF THE INVENTION

[0010] According to an aspect of the invention, the above is at least partly met by a computer-implemented method according to claim 1.

[0011] The general concept of the present invention is based on the fact that it may be possible to improve the classification of objects within a scene by determination of different information about the extracted object, and combine these pieces of information (e.g. a first and a second piece of information about the object as indicated above) based on a probability level for each of the pieces of information. This will advantageously reduce any short term "noise" aspects that may result in error e.g. one of the pieces of information. However, when seen over time and when taking into account the reliability for each of the pieces of information a robust object classification may be achieved.

[0012] This is specifically useful when performing object classification in a "mixed object" scene, e.g. where both a pedestrian and a vehicle are present. This will of course be further advantageous in scenes comprises a plurality of pedestrians and vehicles of different forms and shapes, e.g. including cars, trucks and bikes/bicycles. As a comparison, the above cited US20130182114 is specifically targeting a pre-calibrated environment, i. e. being a typically no-mixed environment where only a single type of object is to be detected and subsequently tracked.

[0013] It may also be possible to further improve the determination of the type of object by including at least a third probability for the type of the object, wherein the third probability is determined by matching the extracted foreground object with a set of predefined images for the plurality of different types of objects, and determining a probability level, being a representation for the third probability, based on a resulting matching level.

[0014] It is advantageous, and within the scope of the invention, to store, for each of the plurality of images, data representing the type of the extracted object(s) and its relative position. Once the data is stored, e.g. within a database, it may for example be possible to track the objects within subsequent images of the image stream. It may also be possible to determine movement pattern for the different types of objects within the scene.

[0015] In the most general implementation of the invention, image data is received and processed. However, it is be possible to further control at least two cameras for acquiring the image stream, wherein each of the at least to cameras are arranged to capturing a predetermined spatially separated corresponding image of the scene. Each image pair will then need to be processed for determining the depth information based on the corresponding images captured by the at least two cameras, for example using a known stereo/3D processing algorithm.

[0016] In some implementations of the invention it may be advantageous to determine the absolute location of the object(s) within the scene. This may for example be achieve by means of an absolute location determined within the scene, where the relative locations for the objects based may be transformed to absolute positions based on the absolute position within the scene. In a possible embodiment, the absolute position within the scene may be acquired from a GPS receiver arranged within the scene.

[0017] According to another aspect of the invention there is provided an image processing system according to claim 10.

[0018] This aspect of the invention provides similar advantages as discussed above in relation to the previous aspects of the invention.

[0019] In a preferred embodiment, the system further comprises at least two cameras wherein each of the at least to cameras are arranged to capturing a predetermined spatially separated corresponding image of the scene. In such an embodiment, the control unit thus further configured to control at least two cameras for acquiring the image stream, and to determine the depth information based on the corresponding images captured by the at least two cameras.

[0020] Within the context of the invention it may alternatively be possible to include a time-of-flight (ToF) camera for acquiring the depth information as to the objects within the scene. In a possible implementation of the invention, a combination of "normal" 2D cameras are used together with at least one ToF camera

[0021] According to a still further aspect of the invention there is provided a computer program product according to claim 13.

[0022] Also this aspect of the invention provides similar advantages as discussed above in relation to the previous aspects of the invention.

[0023] The control unit preferably including a microprocessor or any other type of computing device. Similarly, a software executed by the control unit for operating the inventive wireless communication device may be stored on a computer readable medium, being any type of memory device, including one of a removable nonvolatile random access memory, a hard disk drive, a floppy disk, a CD-ROM, a DVD-ROM, a USB memory, an SD memory card, or a similar computer readable medium known in the art. Accordingly, operation of the wireless communication device may be at least partly automated, implemented as e.g. software, hardware and a combination thereof.

[0024] Further features of, and advantages with, the present invention will become apparent when studying the appended claims and the following description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] The various aspects of the invention, including its particular features and advantages, will be readily understood from the following detailed description and the accompanying drawings, in which:

Fig. 1 illustrates an image processing system according to the invention arranged for collecting and processing image data of a scene;

Fig. 2 provides a conceptual illustration of the image processing system illustrated in Fig. 1;

Fig. 3 is a flow chart illustrating the exemplary steps for operating the system according to the invention, and

Fig. 4 depicts details of object classification according to the invention.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

[0026] The present invention will now be described more fully hereinafter with reference to the accompanying drawings, in which currently preferred embodiments of the invention are shown. This invention may, however, be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided for thoroughness and completeness, and fully convey the scope of the invention to the skilled addressee. Like reference characters refer to like elements throughout.

[0027] Referring now to the drawings and Fig. 1 in particular, there is depicted an image processing system 100 arranged in an elevated position as to a ground/street level, in the illustrated embodiment affixed to an elongated light post 102. The elevated position of the image processing system 100 provides for the possibility of monitoring a scene comprising a plurality of different physical objects. Specifically, in the disclosed embodiment the scene comprises a plurality of different types of physical foreground objects including pedestrians 104, different types of vehicles, such as a truck 106 and a car 108, as well as a plurality of bicycles 110. The scene also comprises a background, for example including a street 112, trees 114, etc. It should be noted that the image processing system 100 may be affixed to or integrated into any form of elevated structure. As such, it could be possible to arrange the image processing system 100 onto e.g. a bridge, an overpass, etc. The image processing system 100 could possibly be integrated with a sign post or any other form of arrangement. The image processing system 100 could be stationary of mobile, connected to the main grid or battery powered.

[0028] The truck 106 and the car 108 travel along the street 112, in opposite directions in different lanes, whereas the pedestrians 104 and the bicycles 110 shares the pavement space at the outsides of the street traveling at mixed directions at each of the pavements.

[0029] In the illustrated embodiment, the image processing system 100 collects image data for a preselected portion of the scene. The size of the preselected portion of the scene may be dependent on the capacity of the image processing system 100, the resolution of the cameras, etc. The collection of image data and possible image processing performed by the image processing system 100 may as exemplified in Fig. 1 be communicated, e.g. wirelessly or wired, to a remote server (not shown), for example over the Internet. Any type of wired or wireless communication protocol could be possible in accordance to the invention.

[0030] Turning now to Fig. 2, which conceptually illustrates the image processing system 100 as shown in Fig. 1. The image processing system 100 comprises a first (left) 202 and a second (right) camera 204 having optical lenses connected thereto, respectively. The image processing system 100 comprises a casing 206, preferably configured for handling outdoor use, the casing holding the cameras 202, 204 as well as processing and communication functionality for the image processing system 100. Typically, the image processing system 100 comprises a control unit 208 arranged in communication with the cameras 202, 204 as well as a communication module 210 (for allowing communication with e.g. the above mentioned remote server) and a memory element 212 (e.g. for intermediate storage of image data). It should be understood that the cameras 202, 204 may not necessarily be provided as separate elements, but may be integrated as one single unit provided with a first and a second set of optical lenses arranged to a first and a second image sensor (i.e. provided for the cameras 202, 204 respectively).

[0031] As understood from the above, the image processing system 100 is configured for stereo collection of image data using the first and the second camera 202, 204. To improve the possibility of creating useful three dimensional image data it is desirable to separate the first and the second camera 202, 204 with a predetermined distance, the distance being dependent on the desired implementation. Furthermore, within the scope of the invention, it should be understood that it could be possible to use other forms of cameras for image collection, such as using a time-of-flight (ToF) camera. Also, the general concept of the invention may be implemented based on image data captured using "external" cameras (i.e. not necessarily including the cameras themselves). That is, the invention may for example be implemented as a server solution where the server receives image data from e.g. remotely arranged cameras. As such, at least some of the processing performed could be implemented in a distributed manner, e.g. partly as a "cloud" solution.

[0032] During operation of the image processing system 100, with further reference to Fig. 3, the process starts by the control unit 208 receiving, S1, an image stream from the first and the second camera 202, 204. The control unit 208 subsequently process, S2, the image stream for forming a three dimensional representation of the scene monitored by the cameras 202, 204. The three dimensional representation is typically formed based on depth information extracted based on the predetermined separation of the first 202 and the second 204 camera, thereby forming depth maps for each image of the image stream (i.e. pair of images from the cameras 202, 204).

[0033] Based on the depth map and a predetermined model of the background it is possible to form, S3, height maps for each of the image pair. The background model defines an "empty" scene where no foreground objects are present. The background model could also be determined in other manners as would be readily understood by the skilled addressee.

[0034] The height maps are subsequently used for extracting a physical foreground object from the scene, such as a pedestrian 104, a car 106 or a bicycle 110. The extracted foreground objects are typically defined to have a "boundary box" closely surrounding the object. The boundary box may typically be rectangular but may also be defined in a more freeform manner.

[0035] For each of the image pair a relative position for the foreground object is determined, S5. The relative position is defined as being relative to e.g. one or a plurality of predetermined "points" within the scene. However, it may of course be possible to make the position "absolute" by correlating one or a plurality of points within the scene to e.g. be a real GPS position, or similar.

[0036] For each extracted object within the scene a first probability as to the type of object is determined, S6. In a preferred embodiment of the invention a set of object types are predefined. Each of the set of the predefined object types are defined to have at least a length, a height and a width being within a predetermined range. The foreground object being extracted from the scene is accordingly matched to the set of predefined object types, and a probability is determined for the extracted foreground object as compared to the predefined object types. As understood, the set of predefined object types typically includes a set of height, width and length parameters defining the pedestrian, the bicycle, the car and the truck. The size (e.g. the boundary box as mentioned above) of the extracted foreground object is thus compared to the predefined parameters for each of the type of objects for determining the first probability.

[0037] A second probability is determined, S7, where the extracted foreground object essentially is "tracked" between subsequent image pairs of the image stream. In accordance to the invention a difference in the relative position is determined for the extracted foreground object, i.e. from one image pair to the next. Accordingly, it may be possible to, for each of the predefined object types, form a set of "speed" parameters/profile defining how fast the specific object type likely will move between subsequent image pairs. The tracking will generally also include determination of the direction of movement of the extracted object.

[0038] Finally, the type of the foreground object is determined, S8, by combining the first and the second probability, possibly with further information for increasing the probability of correctly determining the object type for the extracted foreground object.

[0039] This is further exemplified in Fig. 4 showing details of the object classification method according to the invention. Specifically, the flowchart illustrates the above discussed geometric matching classification, e.g. including matching of size and "moving object positions" for subsequent image pairs with predefined object type parameters (i.e. 404 and 406). In addition, a further, e.g. possibly defined as a third probability may be determined by performing an image matching 408 between the extracted foreground object and representative image data stored in a database 410. The image matching may preferably take into account the direction of movement of the extracted object, for example for use in improving the selection of the "correct" image to match with, i.e. as the stored images to match with may be annotated to include information as to an aspect angle for the stored image. The third probability for the object type may typically be combined with the first and the second probability.

[0040] It should be understood that the combination of the first, second and third probability may be weighted, e.g. more weight could for example be given to the height, width and length parameter matching as compared to the remaining two probabilities. Any type of weighting may be possible and is within the scope of the invention. Also, it should be understood that the combined probability may be accumulated over time, i.e. by allowing further geometric matching classifications and/or image based classifications to be made for the same object. That is, it will likely be preferred to continue to track the object as long as the object is visible within the scene. Thereby, the likelihood of determining the correct object type will be increased.

[0041] Once the object type has been determined, i.e. being one of the predefined types of object, this knowledge may be used for a numerous different applications. For example, the image processing system 100 could be used in a traffic counting scenario comprising tracking mixed object types for categorizing the behavior for each of the object types. Similarly, the image processing system 100 according to the invention could be used for controlling traffic patterns, i.e. to control traffic lights in a city, for example by adapting time given for passing a street. For example, at peak hours with a high density of bicycles, the bicycles could be given more passing time as compared to cars, possibly improving the overall traffic environment within the city.

[0042] In addition, the control functionality of the present disclosure may be implemented using existing computer processors, or by a special purpose computer processor for an appropriate system, incorporated for this or another purpose, or by a hardwired system. Embodiments within the scope of the present disclosure include program products comprising machine-readable media for carrying or having machine-executable instructions or data structures stored thereon. Such machine-readable media can be any available media that can be accessed by a general purpose or special purpose computer or other machine with a processor. By way of example, such machine-readable media can comprise RAM, ROM, EPROM, EEPROM, CD-ROM or other optical disk storage, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to carry or store desired program code in the form of machine-executable instructions or data structures and which can be accessed by a general purpose or special purpose computer or other machine with a processor. When information is transferred or provided over a network or another communications connection (either hardwired, wireless, or a combination of hardwired or wireless) to a machine, the machine properly views the connection as a machine-readable medium. Thus, any such connection is properly termed a machine-readable medium. Combinations of the above are also included within the scope of machine-readable media. Machine-executable instructions include, for example, instructions and data which cause a general purpose computer, special purpose computer, or special purpose processing machines to perform a certain function or group of functions.

Claims

1. A computer-implemented method for classifying a type of a moving object (104, 106, 108, 110) within a scene, the moving object being one of a plurality of different types of predetermined objects, the method comprising:

- receiving (S1) an image stream comprising a plurality of images from a first (202) and a second (204) camera arranged in an elevated position as to a ground level within the scene, wherein each of the first and the second camera

is arranged to capture a predetermined spatially separated corresponding image of the scene together forming a pair of images, each of the pair of images being a captured representation of the scene, wherein the content of the scene comprises at least one physical foreground object and a physical background;
- processing (S2) the pair of images to produce, for each pair of images, a depth map;
- forming (S3) a height map for each pair of images using a predetermined model of the background and the corresponding depth map, the background model defining the scene when no physical foreground objects are present;
- extracting (S4) the at least one physical foreground object from each pair of images using the corresponding height map;
- determining (S5), for each pair of images, a relative position of the at least one physical foreground object within the scene using the height map, the relative position being relative to a plurality of predetermined points within the scene;
- determining (S6) a first probability level for the type of the at least one foreground object within each pair of images by matching the at least one extracted physical foreground object with a predetermined set of different object types, wherein each object type within the set has a length, a height and a width being within a predetermined range;
- determining (S7) a second probability level for the type of the at least one foreground object by determining a difference in the relative position of the at least one foreground object in at least two images of the plurality of images, converting the difference in relative position to a relative speed for the at least one foreground object, and matching the relative speed with speed profiles for the predetermined set of different object types; and
- combining (S8) the first and the second probability level to define the type of the at least one foreground object.

2. The method according to claim 1, wherein the different types of predetermined objects comprise a pedestrian, a bicycle and a vehicle.

3. The method according to any one of claims 1 and 2, further defining the type of the at least one foreground object based on a third probability level for the type of the at least one foreground object, wherein the third probability is determined by:

- matching the at least one extracted foreground object with a set of predefined images for the plurality of different types of objects, and
- determining a probability level, being a representation for the third probability, based on a resulting matching level.

4. The method according to any one of claims 1 and 2, wherein the determination of the first and the second probability level is defined to be formed by a geometrical classifier.

5. The method according to claim 3, wherein the determination of the third probability level is defined to be formed by an image-based classifier.

6. The method according to any one of claims 1 and 2, further comprising:

- determining a three-dimensional boundary box surrounding the at least one extracted foreground object.

7. The method according to any one of claims 1 and 2, further comprising:

- tracking the at least one extracted foreground object within subsequent images of the image stream.

8. The method according to any one of the preceding claims, further comprising:

- storing, for each of the plurality of images, data representing the type of the at least one extracted foreground object and its relative position within a database.

9. The method according to any one of the preceding claims, further comprising:

- determining an absolute position for the at least one extracted foreground object based on an absolute position for at least one location within the scene.

10. An image processing system (100) for classifying a type of a moving object (104, 106, 108, 110) within a scene, the moving object being one of a plurality of different types of predetermined objects, the image processing system comprising a control unit configured to:

- receive an image stream comprising a plurality of images from a first (202) and a second (204) camera arranged in an elevated position as to a ground level within the scene, wherein each of the first and the second camera is arranged to capture a predetermined spatially separated corresponding image of the scene together forming a pair of images, each of the pair of images being a captured representation of the

scene, wherein the content of the scene comprises at least one physical foreground object and a physical background;
- process the plurality of images to produce, for each pair of images, a depth map;
- form a height map for each pair of images using a predetermined model of the background and the corresponding depth map;
- extract the at least one physical foreground object from each of pair of images using the corresponding height map;
- determine, for each of the pair of images, a relative position of the at least one physical foreground object within the scene using the height map, the background model defining the scene when no physical foreground objects are present;
- determine a first probability level for the type of the at least one foreground object within each of the images by matching the at least one extracted physical foreground object with a predetermined set of different object types, wherein each object type within the set has a length, a height and a width being within a predetermined range;
- determine a second probability level for the type of the at least one foreground object by determining a difference in the relative position for the at least one foreground object in at least two images of the plurality of images, converting the difference in relative position to a relative speed for the at least one foreground object, and matching the relative speed with speed profiles for the predetermined set of different object types; and
- combine the first and the second probability level to define the type of the at least one foreground object.

11. The system according to claim 10, further comprising:

- a position interface for receiving an absolute position for at least one location within the scene,

wherein the control unit is further configured to:

- determine an absolute position for the at least one extracted foreground object based on the absolute position.

12. The system according to claim 11, further comprising a GPS receiver connected to the position interface and providing the absolute position of the system.

13. A computer program product comprising instructions, which when the program is executed by a computer, cause the computer to carry out the method of claim 1.

Patentansprüche

1. Computerimplementiertes Verfahren zum Klassifizieren eines Typs eines sich bewegenden Objekts (104, 106, 108, 110) innerhalb einer Szene, wobei das sich bewegende Objekt eines von mehreren verschiedenen Typen von vorbestimmten Objekten ist, wobei das Verfahren umfasst:

- Empfangen (S1) eines Bildstroms, der mehrere Bilder umfasst, von einer ersten (202) und einer zweiten (204) Kamera, die in einer erhöhten Position in Bezug auf eine Bodenebene innerhalb der Szene angeordnet sind, wobei die erste Kamera und die zweite Kamera derart angeordnet sind, dass eine jeweilige Kamera ein vorbestimmtes, räumlich getrenntes, entsprechendes Bild der Szene aufnimmt, wobei die aufgenommenen Bilder zusammen ein Paar von Bildern bilden, wobei jedes Bild des Paars von Bildern eine aufgenommene Darstellung der Szene ist, wobei der Inhalt der Szene mindestens ein physisches Vordergrundobjekt und einen physischen Hintergrund umfasst;
- Verarbeiten (S2) des Paars von Bildern, um für jedes Paar von Bildern eine Tiefenkarte zu erzeugen;
- Bilden (S3) einer Höhenkarte für jedes Paar von Bildern unter Verwendung eines vorbestimmten Modells des Hintergrunds und der entsprechenden Tiefenkarte, wobei das Hintergrundmodell die Szene definiert, wenn keine physischen Vordergrundobjekte vorhanden sind;
- Extrahieren (S4) des mindestens einen physischen Vordergrundobjekts aus jedem Paar von Bildern unter Verwendung der entsprechenden Höhenkarte;
- Bestimmen (S5), für jedes Paar von Bildern, einer relativen Position des mindestens einen physischen Vordergrundobjekts innerhalb der Szene unter Verwendung der Höhenkarte, wobei die relative Position relativ zu mehreren vorbestimmten Punkten innerhalb der Szene ist;
- Bestimmen (S6) eines ersten Wahrscheinlichkeitsniveaus für den Typ des mindestens einen Vordergrundobjekts innerhalb jedes Paars von Bildern durch Abgleichen des mindestens einen extrahierten physischen Vordergrundobjekts mit einem vorbestimmten Satz von verschiedenen Objekttypen, wobei jeder Objekttyp innerhalb des Satzes eine Länge, eine Höhe und eine Breite aufweist, die innerhalb eines vorbestimmten Bereichs liegen;

- Bestimmen (S7) eines zweiten Wahrscheinlichkeitsniveaus für den Typ des mindestens einen Vordergrundobjekts durch Bestimmen einer Differenz zwischen den relativen Positionen des mindestens einen Vordergrundobjekts in mindestens zwei Bildern der mehreren Bilder, Umwandeln der Differenz der relativen Positionen in eine relative Geschwindigkeit für das mindestens eine Vordergrundobjekt und Abgleichen der relativen Geschwindigkeit mit Geschwindigkeitsprofilen für den vorbestimmten Satz von verschiedenen Objekttypen; und
- Kombinieren (S8) des ersten und des zweiten Wahrscheinlichkeitsniveaus, um den Typ des mindestens einen Vordergrundobjekts zu definieren.

2. Verfahren nach Anspruch 1, wobei die verschiedenen Typen von vorbestimmten Objekten einen Fußgänger, ein Fahrrad und ein Fahrzeug umfassen.

3. Verfahren nach einem der Ansprüche 1 und 2, ferner umfassend Definieren des Typs des mindestens einen Vordergrundobjekts basierend auf einem dritten Wahrscheinlichkeitsniveau für den Typ des mindestens einen Vordergrundobjekts, wobei die dritte Wahrscheinlichkeit bestimmt wird durch:

- Abgleichen des mindestens einen extrahierten Vordergrundobjekts mit einem Satz von vordefinierten Bildern für die mehreren verschiedenen Typen von Objekten, und
- Bestimmen eines Wahrscheinlichkeitsniveaus, das eine Darstellung für die dritte Wahrscheinlichkeit ist, basierend auf einem resultierenden Abgleichniveau.

4. Verfahren nach einem der Ansprüche 1 und 2, wobei die Bestimmung des ersten und des zweiten Wahrscheinlichkeitsniveaus definiert ist, durch einen geometrischen Klassifikator gebildet zu werden.

5. Verfahren nach Anspruch 3, wobei die Bestimmung des dritten Wahrscheinlichkeitsniveaus definiert ist, durch einen bildbasierten Klassifikator gebildet zu werden.

6. Verfahren nach einem der Ansprüche 1 und 2, ferner umfassend:

- Bestimmen eines dreidimensionalen Begrenzungsrahmens, der das mindestens eine extrahierte Vordergrundobjekt umgibt.

7. Verfahren nach einem der Ansprüche 1 und 2, ferner umfassend:

- Verfolgen des mindestens einen extrahierten Vordergrundobjekts innerhalb aufeinanderfolgender Bilder des Bildstroms.

8. Verfahren nach einem der vorhergehenden Ansprüche, ferner umfassend:

- Speichern, für jedes der mehreren Bilder, von Daten, die den Typ des mindestens einen extrahierten Vordergrundobjekts und seine relative Position darstellen, in einer Datenbank.

9. Verfahren nach einem der vorhergehenden Ansprüche, ferner umfassend:

- Bestimmen einer absoluten Position für das mindestens eine extrahierte Vordergrundobjekt basierend auf einer absoluten Position für mindestens einen Ort in der Szene.

10. Bildverarbeitungssystem (100) zum Klassifizieren eines Typs eines sich bewegenden Objekts (104, 106, 108, 110) innerhalb einer Szene, wobei das sich bewegende Objekt ein Objekt eines Typs aus mehreren verschiedenen Typen von vorbestimmten Objekten ist, wobei das Bildverarbeitungssystem eine Steuereinheit aufweist, die eingerichtet ist zum:

- Empfangen eines Bildstroms, der mehrere Bilder umfasst, von einer ersten (202) und einer zweiten (204) Kamera, die in einer erhöhten Position in Bezug auf eine Bodenebene innerhalb der Szene angeordnet sind, wobei die erste Kamera und die zweite Kamera derart angeordnet sind, dass eine jeweilige Kamera ein vorbestimmtes, räumlich getrenntes, entsprechendes Bild der Szene aufnimmt, wobei die aufgenommenen Bilder zusammen ein Paar von Bildern bilden, wobei jedes Bild des Paars von Bildern eine aufgenommene Darstellung der Szene ist, wobei der Inhalt der Szene mindestens ein physisches Vordergrundobjekt und einen physischen Hintergrund umfasst;
- Verarbeiten der mehreren Bilder, um für jedes Paar von Bildern eine Tiefenkarte zu erzeugen;
- Bilden einer Höhenkarte für jedes Paar von Bildern unter Verwendung eines vorbestimmten Modells des Hintergrunds und der entsprechenden Tiefenkarte;
- Extrahieren des mindestens einen physischen Vordergrundobjekts aus jedem Paar von Bildern unter Verwendung der entsprechenden Höhenkarte;
- Bestimmen, für jedes Paar von Bildern, einer relativen Position des mindestens einen physischen Vordergrundobjekts innerhalb der Szene unter Verwendung der Höhenkarte, wobei das Hintergrundmodell die Szene definiert, wenn

keine physischen Vordergrundobjekte vorhanden sind;

- Bestimmen eines ersten Wahrscheinlichkeitsniveaus für den Typ des mindestens einen Vordergrundobjekts innerhalb jedes der Bilder durch Abgleichen des mindestens einen extrahierten physischen Vordergrundobjekts mit einem vorbestimmten Satz von verschiedenen Objekttypen, wobei jeder Objekttyp innerhalb des Satzes eine Länge, eine Höhe und eine Breite aufweist, die innerhalb eines vorbestimmten Bereichs liegen;
- Bestimmen eines zweiten Wahrscheinlichkeitsniveaus für den Typ des mindestens einen Vordergrundobjekts durch Bestimmen einer Differenz zwischen den relativen Positionen des mindestens einen Vordergrundobjekts in mindestens zwei Bildern der mehreren Bilder, Umwandeln der Differenz der relativen Positionen in eine relative Geschwindigkeit für das mindestens eine Vordergrundobjekt und Abgleichen der relativen Geschwindigkeit mit Geschwindigkeitsprofilen für den vorbestimmten Satz von verschiedenen Objekttypen; und
- Kombinieren des ersten und des zweiten Wahrscheinlichkeitsniveaus, um den Typ des mindestens einen Vordergrundobjekts zu definieren.

11. System nach Anspruch 10, ferner umfassend:

- eine Positionsschnittstelle zum Empfangen einer absoluten Position für mindestens einen Ort innerhalb der Szene,

wobei die Steuereinheit ferner eingerichtet ist zum:

- Bestimmen einer absoluten Position für das mindestens eine extrahierte Vordergrundobjekt basierend auf der absoluten Position.

12. System nach Anspruch 11, ferner umfassend einen GPS-Empfänger, der mit der Positionsschnittstelle verbunden ist und die absolute Position des Systems bereitstellt.

13. Computerprogrammprodukt, umfassend Anweisungen, die, wenn das Programm von einem Computer ausgeführt wird, den Computer veranlassen, das Verfahren nach Anspruch 1 auszuführen.

Revendications

1. Procédé mis en œuvre par ordinateur pour classer un type d'objet en mouvement (104, 106, 108, 110) dans une scène, l'objet en mouvement étant un d'une pluralité de différents types d'objets prédéterminés, le procédé comprenant de :

- recevoir (S1) un flux d'images comprenant une pluralité d'images d'une première (202) et une deuxième (204) caméra agencée dans une position surélevée par rapport au niveau du sol dans la scène, dans laquelle chacune de la première et de la deuxième caméra est agencée pour capturer une image correspondante prédéterminée spatialement séparée de la scène formant ensemble une paire d'images, chacune de la paire d'images étant une représentation capturée de la scène, dans lequel le contenu de la scène comprend au moins un objet physique de premier plan et un arrière-plan physique ;
- traiter (S2) le couple d'images pour produire, pour chaque couple d'images, une carte de profondeur;
- former (S3) une carte de hauteur pour chaque paire d'images en utilisant un modèle prédéterminé d'arrière-plan et la carte de profondeur correspondante, le modèle d'arrière-plan définissant la scène lorsqu'aucun objet physique de premier plan n'est présent ;
- extraire (S4) le au moins un objet physique de premier plan de chaque paire d'images en utilisant la carte de hauteur correspondante ;
- déterminer (S5), pour une paire d'images, une position relative dudit au moins un objet physique de premier plan dans la scène en utilisant la carte de hauteur, la position relative étant relative à une pluralité de points prédéterminés dans la scène ;
- déterminer (S6) un premier niveau de probabilité pour le type d'au moins un objet de premier plan dans chaque paire d'images en faisant concorder le au moins un objet de premier plan physique extrait avec un ensemble prédéterminé de types d'objets différents, dans lequel chaque type d'objet dans l'ensemble a une longueur, une hauteur et une largeur comprises dans une plage prédéterminée ;
- déterminer (S7) un deuxième niveau de probabilité pour le type du au moins un objet de premier plan en déterminant une différence dans la position relative du au moins un objet de premier plan dans au moins deux images de la pluralité d'images, la conversion de la différence de position relative en une vitesse relative pour l'objet d'au moins un objet de premier plan, et la mise en concordance de la vitesse relative avec des profils de vitesse pour l'ensemble prédéterminé de différents types d'objets ; et
- combiner (S8) le premier et le deuxième niveau de probabilité pour définir le type d'au moins un objet de premier plan.

2. Procédé selon la revendication 1, dans lequel les différents types d'objets prédéterminés comprennent un piéton, un vélo et un véhicule.

3. Procédé selon une quelconque des revendications 1 et 2, définissant en outre le type d'au moins un objet de premier plan sur la base d'un troisième niveau de probabilité pour le type d'au moins un objet de premier plan, dans lequel la troisième probabilité est déterminée par :

- faire concorder le au moins un objet de premier plan extrait avec un ensemble d'images prédéfinies pour la pluralité de différents types d'objets, et
- déterminer un niveau de probabilité, étant une représentation pour la troisième probabilité, sur la base d'un niveau de concordance résultant.

4. Procédé selon une quelconque des revendications 1 et 2, dans lequel la détermination du premier et du deuxième niveau de probabilité est définie pour être formée par un classificateur géométrique.

5. Procédé selon la revendication 3, dans lequel la détermination du troisième niveau de probabilité est définie pour être formée par un classificateur à base d'images.

6. Procédé selon une quelconque des revendications 1 et 2, comprenant en outre de :

- déterminer une zone de contour tridimensionnelle entourant le au moins un objet de premier plan extrait.

7. Procédé selon une quelconque des revendications 1 et 2, comprenant en outre de :

- suivre le au moins un objet de premier plan extrait dans les images suivantes du flux d'images.

8. Procédé selon une quelconque des revendications précédentes, comprenant en outre de :

- mémoriser, pour chacune de la pluralité d'images, des données représentant le type d'au moins un objet de premier plan extrait et sa position relative dans une base de données.

9. Procédé selon une quelconque des revendications précédentes, comprenant en outre de :

- déterminer une position absolue pour le au moins un objet de premier plan extrait sur la base d'une position absolue pour au moins un emplacement dans la scène.

10. Système de traitement d'image (100) pour classer un type d'objet en mouvement (104, 106, 108, 110) dans une scène, l'objet en mouvement étant un d'une pluralité de différents types d'objets prédéterminés, le système de traitement d'image comprenant une unité de commande configurée pour :

- recevoir un flux d'images comprenant une pluralité d'images provenant d'une première (202) et d'une deuxième (204) caméras agencées dans une position surélevée par rapport au niveau du sol dans la scène, dans lequel chacune des première et deuxième caméras est agencée pour capturer une image correspondante prédéterminée spatialement séparée de la scène formant ensemble une paire d'images, chacune de la paire d'images étant une représentation capturée de la scène, dans lequel le contenu de la scène comprend au moins un objet physique de premier plan et un arrière-plan physique ;
- traiter la pluralité d'images pour produire pour chaque couple d'images, une carte de profondeur;
- former une carte de hauteur pour chaque paire d'images en utilisant un modèle prédéterminé d'arrière-plan et la carte de profondeur correspondante ;
- extraire le au moins un objet physique de premier plan de chacune des paires d'images en utilisant la carte de hauteur correspondante ;
- déterminer, pour chacune des paires d'images, une position relative d'au moins un objet physique de premier plan dans la scène en utilisant la carte de hauteur, le modèle d'arrière-plan définissant la scène lorsqu'aucun objet physique de premier plan n'est présent ;
- déterminer un premier niveau de probabilité pour le type d'au moins un objet de premier plan dans chacune des images en faisant concorder le au moins un objet physique de premier plan extrait avec un ensemble prédéterminé de différents types d'objets, dans lequel chaque type d'objet dans l'ensemble a une longueur, une hauteur et une largeur comprises dans une plage prédéterminée ;
- déterminer un deuxième niveau de probabilité pour le type d'au moins un objet de premier plan en déterminant une différence de position relative pour le au moins un objet de premier plan dans au moins deux images de la pluralité d'images, en convertissant la différence de position relative en une vitesse relative pour le au moins un objet de premier plan, et en faisant concorder la vitesse relative avec des profils de vitesse pour l'ensemble prédéterminé de différents types d'objets ; et
- combiner le premier et le deuxième niveau de probabilité pour définir le type du au moins un

objet de premier plan.

11. Système selon la revendication 10, comprenant en outre de :

- une interface de position pour recevoir une position absolue pour au moins un emplacement dans la scène, dans lequel l'unité de commande est en outre configurée pour :
- déterminer une position absolue pour le au moins un objet de premier plan extrait sur la base de la position absolue.

12. Système selon la revendication 11, comprenant en outre un récepteur GPS relié à l'interface de position et fournissant la position absolue du système.

13. Produit de programme informatique comprenant des instructions qui, lorsque le programme est exécuté par un ordinateur, amènent l'ordinateur à exécuter le procédé de la revendication 1.

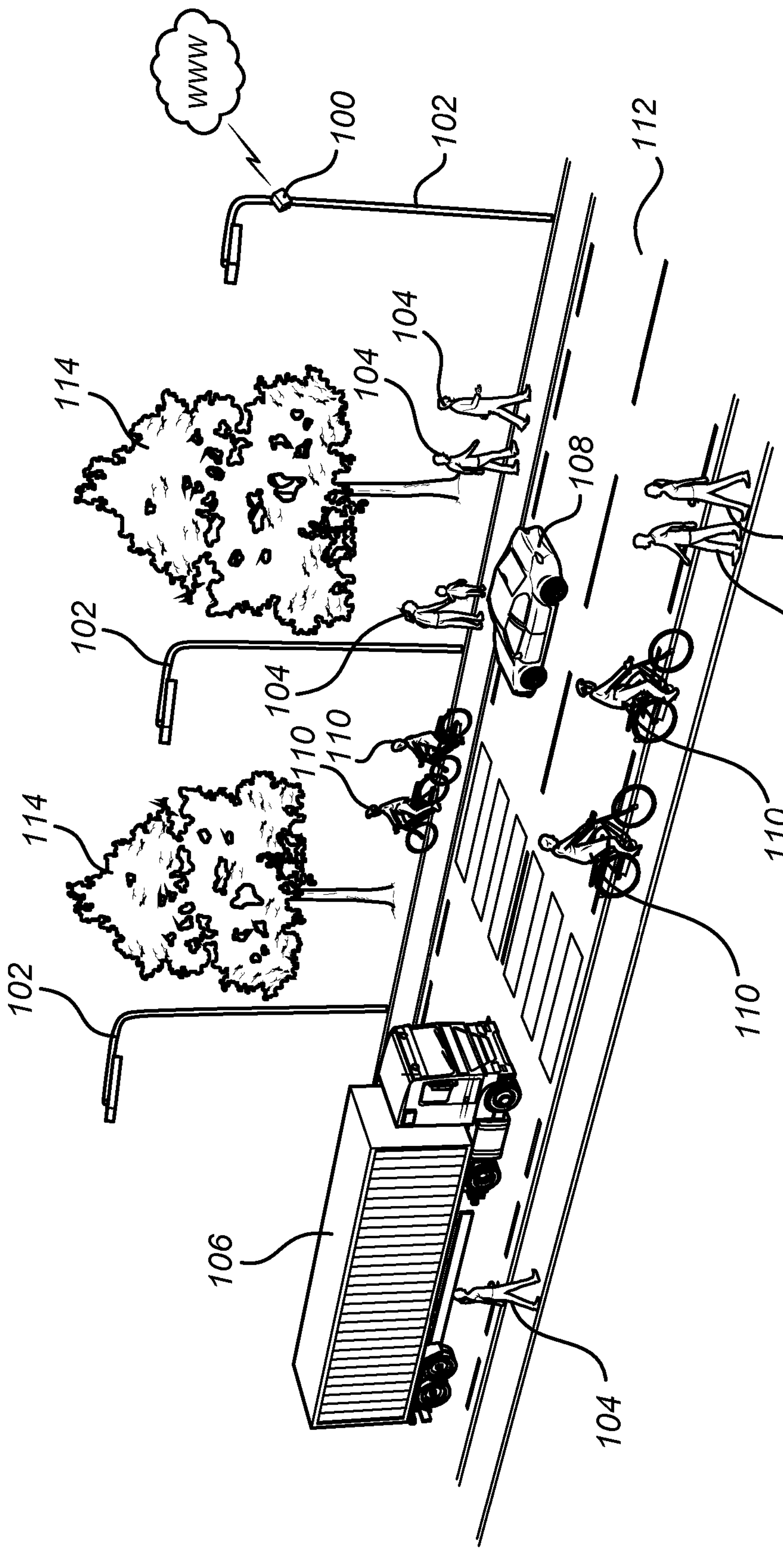

Fig. 1

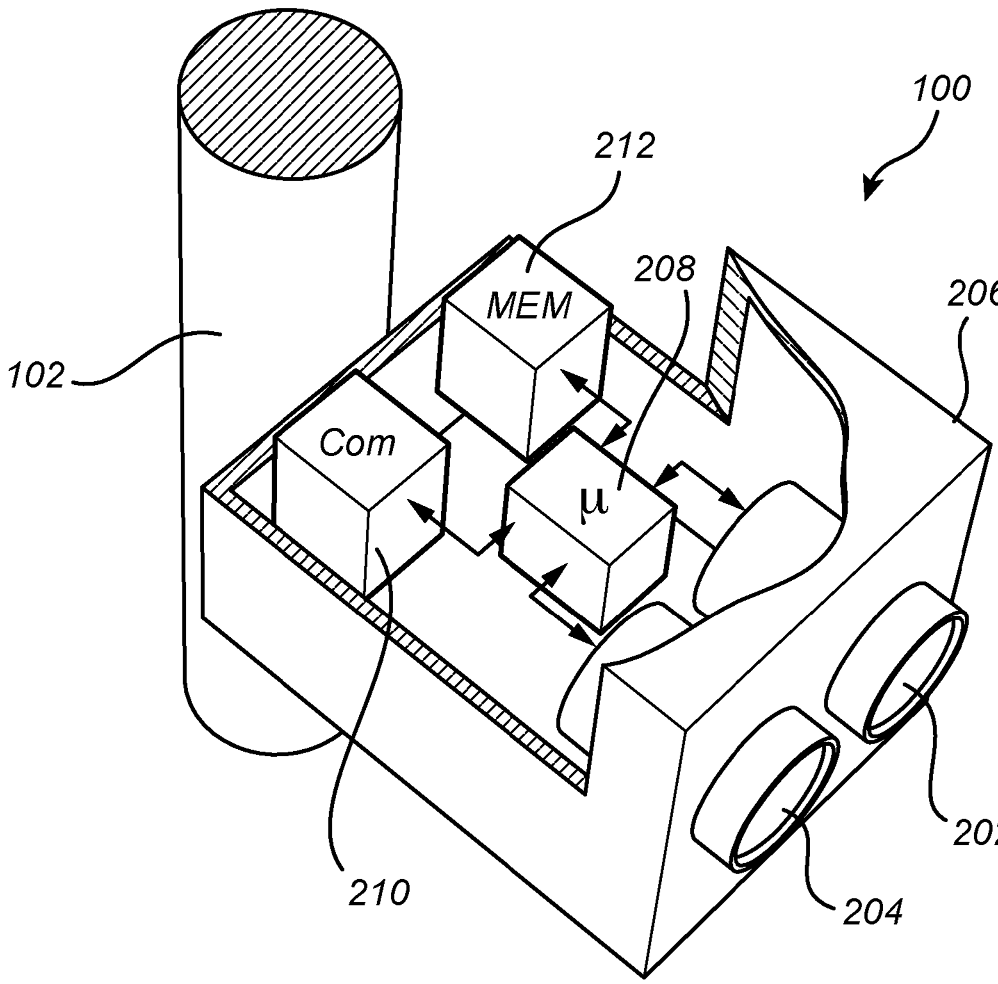

Fig. 2

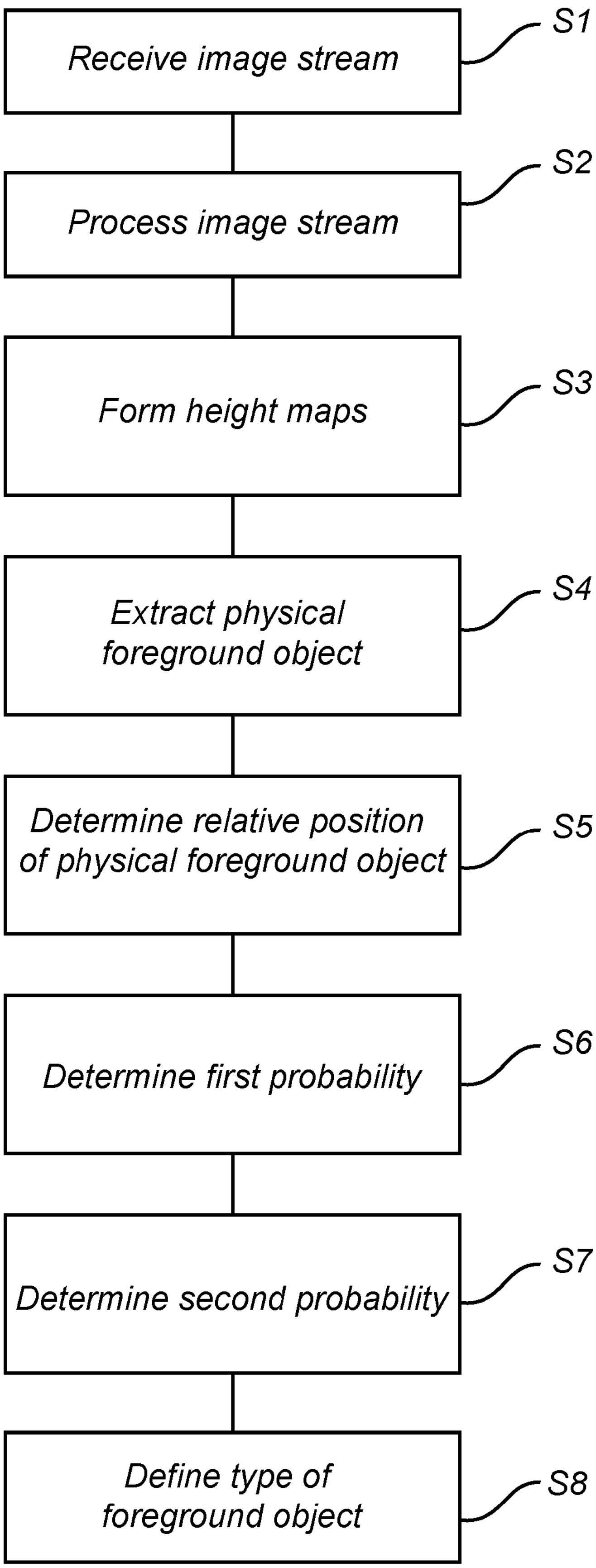
Receive image stream
S1
Process image stream
S2
Form height maps
S3
Extract physical foreground object
S4
Determine relative position of physical foreground object
S5
Determine first probability
S6
Determine second probability
S7
Define type of foreground object
S8

Fig. 3

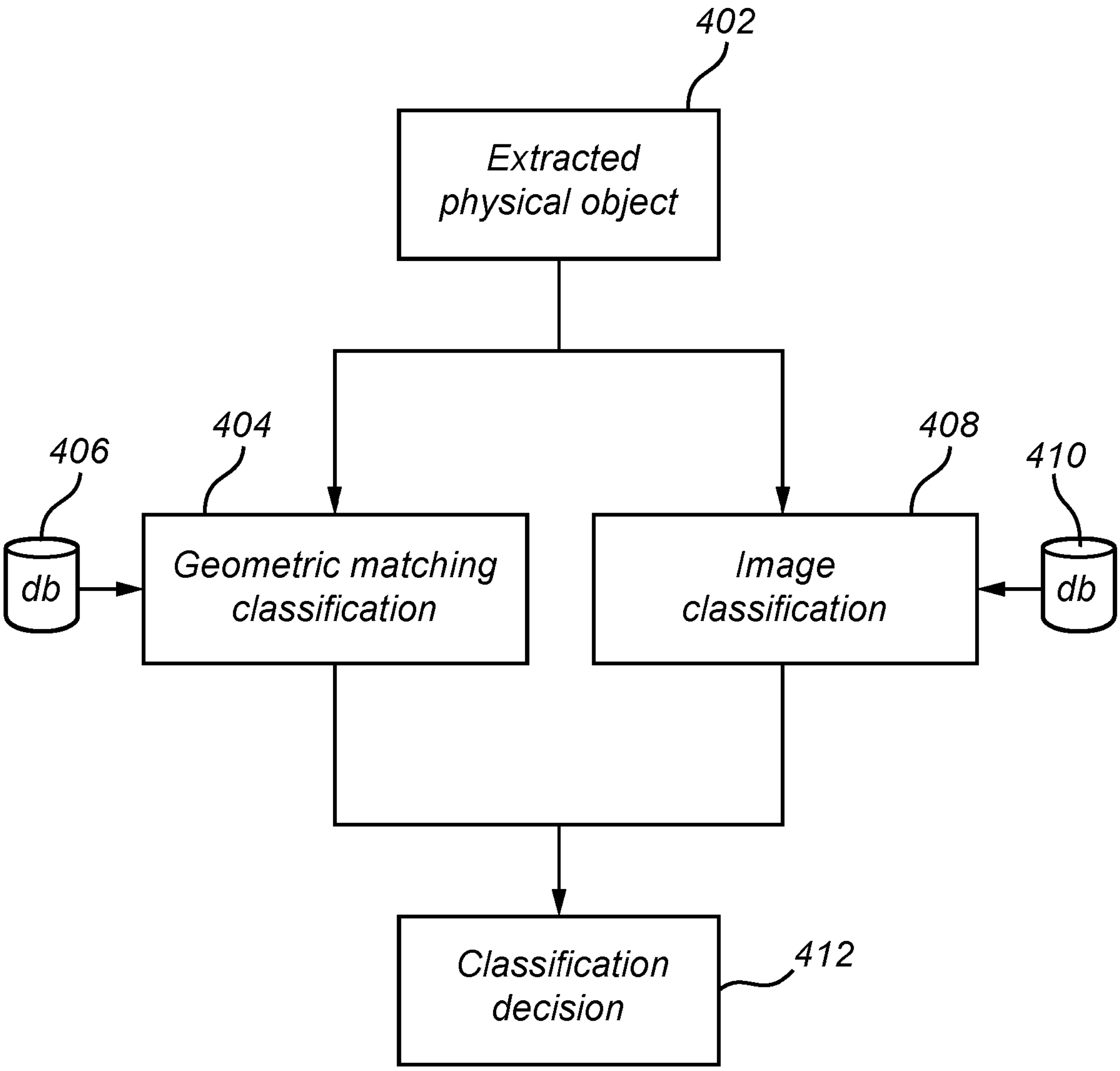
402
Extracted physical object
404
Geometric matching classification
406
db
408
Image classification
410
db
Classification decision
412

Fig. 4

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 7602944 B **[0004] [0005]**
- US 20130182114 A **[0006] [0012]**
- CN 101388145 **[0007]**
- US 20090195371 A **[0009]**
- US 2009010490 A **[0009]**

Non-patent literature cited in the description

- A framework for object detection, tracking and classification in urban traffic scenarios using stereovision. **BOTA S et al.** INTELLIGENT COMPUTER COMMUNICATION AND PROCESSING, 2009. ICCP 2009. IEEE 5TH INTERNATIONAL CONFERENCE ON. IEEE, 27 August 2009, 153-156 **[0008]**